DOCUMENT RESUME

ED 144 966 TM 006 459

AUTHOR Cordray, D. S.; Staneski, R. A.
TITLE A Model for Testing Rival Hypotheses in Longitudinal Social Problems.
PUB DATE 8 Apr 76
NOTE 24p.; Paper presented at the Annual Meeting of the Western Psychological Association (56th, Los Angeles, California, April 8-11, 1976) ; Product of the Head Start Primary Continuation Learning Project ; Tables may be marginally legible due to print quality

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Attrition (Research Studies); *Compensatory Education Programs; Data Analysis; Evaluation Criteria; *Evaluation Methods; Longitudinal Studies; *Models; Primary Education; *Program Evaluation; Research Methodology; *Research Problems; Student Mobility; Validity
IDENTIFIERS Project Head Start

ABSTRACT

Some of the problems unique to evaluation research, in particular to the evaluation of longitudinal studies of social programs, are presented. A working model found to be useful in structuring compensatory education program evaluation activities is also offered. Since the goals of social programs such as Head Start are dynamic rather than static, the evaluation design must be flexible in order to accomodate unexpected events and alterations in the program process. Attrition and the transiency of student program participants are highlighted as two of the chief problem areas in evaluating longitudinal studies. (MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED144966

A Model for Testing Rival Hypotheses in Longitudinal Social Programs

D. S. Cordray and R. A. Staneski
Evaluation Staff: Head Start
Primary Continuation Learning Project
California State University, Northridge
and
Claremont Graduate School

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.S. Cordray
R.A. Staneski

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The work upon which this paper is based was performed pursuant to Contract HEW-100-75-0051 with the Department of Health, Education and Welfare

Presented at the Western Psychological Association meetings, Los Angeles, California April 8, 1976

TM006459

The evaluation of social programs presents problems that are typically not of major concern to a laboratory researcher. While the goals of both the laboratory and the evaluation researcher (i.e., the gathering of accurate information regarding some problem) are much the same, the means by which these are attained may be considerably different. The present paper will discuss some of the problems unique to evaluation research with a particular emphasis on the evaluation of longitudinal programs. Further, we will present a "working model" which we have found useful in structuring our own evaluation activities.

There currently exist numerous excellent references which discuss the problems and methodological issues to be considered before undertaking a program evaluation (e.g., Weiss, 1972; Hyman & Wright, 1967; Tripodi, Fellin & Epstein, 1971). While most of the problems discussed in the above references are typically encountered within evaluative research, longitudinal programs manifest additional problems which should be considered during the planning, implementation, and data gathering stages of evaluation. The distinction between longitudinal and "short run" programs is rather arbitrary and most likely depends on the nature of the problem under consideration. The term "longitudinal" is currently being employed to describe those programs that involve an assessment of the same individuals over an extended period of time. This extended time frame (e.g., one year or more) creates the basis for sources of variation (e.g., attrition, length of exposure to the program, etc.) which are accentuated in contrast with those programs that do not have sustained contact with their recipients.

Having drawn this distinction, it is possible to turn our attention towards some of the problems associated with evaluating long term programs.

It has been our experience that program goals and objectives, no matter how carefully they have been initially specified, are dynamic rather than static. Consequently, changes in program structure over time preclude the use of a rigid a priori evaluation strategy. The evaluation design, especially for longitudinal programs, must be flexible in order to accommodate unexpected events and alterations in the program process and structure. Further, these alterations bring with them increased numbers of alternative explanations for a program's success or failure. Strategies for the identification and examination of these systematic sources of variation provides the basis of the present paper.

To illustrate some of the problems we have encountered, let us begin with a brief description of the structure, evaluation hypothesis, and evaluation design of the particular program with which we are currently associated. We will then attempt to point out what we call "sources of systematic variation" which have been encountered during the course of our evaluation. Finally, we will provide a framework for examining these sources of variation.

Description of Program

The current paper stems from four years of preliminary evaluation of the five-year Head Start Primary Continuation Learning Project (HSPCLP). The program is located in three highly mobile communities in Southern California. While this program is multi-faceted, one of the primary objectives is to provide continuity between Head Start and public schools with the goal of enhancing cognitive and social development. The Head Start experience was provided by the Foundation for Early Childhood Education, Inc. which has a child-development orientation. Preschool aged children entered Foundation Head Start (FHS) in which selected parents

from their communities acted as Assistant Teachers. As FHS Assistant teachers, they received inservice education in child development. Subsequent to one year of FHS experience, the pupils entered public schools with their FHS Assistant Teacher who had two to six years experience in Head Start. The role of the FHS Assistant Teachers was to provide continuity for the pupils between the Foundation Head Start program and their ensuing public school experiences through the third grade. This aspect of the continuation program was designed to sustain the gains effected at Head Start. Further, it was hypothesized that the presence of the Continuation Teacher (C.T.)[1] in HSPCLP classes would provide increased cognitive and social skills for pupils who had not been exposed to the FHS experience in comparison to pupils enrolled in a traditional classroom setting. The HSPCLP includes two waves of Continuation Teachers and pupils. Wave I entered kindergarten in Fall, 1971, and Wave II entered kindergarten one year later (Fall, 1972). In terms of an experimental design, Wave II can be viewed as a replication of Wave I.

Evaluation Considerations

After reviewing the prototypical quasi-designs provided by Campbell & Stanley (1963) and Cook & Campbell (1976) it becomes obvious that it is not always possible to apply any particular design in a given research situation. However, these designs are merely examples of the means by which an evaluator attempts to ascertain the validity of his or her research findings. Whether or not these prototypes can be applied to any given situation is secondary to the concept of testing rival hypotheses,

---

[1] FHS Assistant Teachers were called Continuation Teachers while involved in the Continuation Project.

which is the essence of quasi-experimentation. However, the validity of a finding is questionable primarily when a treatment effect has been demonstrated to be statistically reliable. In addition to the testing of rival hypotheses, when program effects are present, it is of equal importance to isolate sources of variation which result in the masking of program effects. This latter problem is especially important when the treatment is weak and implementation of the program must be maintained over long periods of time. The determination of program effects must be monitored at various intervals throughout its duration if summative evaluation is going to be able to specify whether or not the program, as planned, obtained its initial objectives.

The evaluation process can be viewed as consisting of two stages, the first of these being the isolation of systematic variation within the project and then an assessment of alternative explanations. The second stage (espoused by Campbell, et al) would involve the elimination of rival hypotheses. Since the procedure for testing rival hypotheses has been outlined elsewhere (Campbell & Stanley, 1963 and Cook & Campbell, 1976) our attention will be focused on the assessment of suppressor variables which could lead to erroneous conclusions regarding the program's effects.

Problematic Sources of Variation

Undesirable systematic variation occurs in any logitudinal program. For instance, our experience has shown that variation exists both within and between project sites, the sources of which are not always obvious. A related problem results from the non-random assignment of implementation sites, recipients, and personnel within the structure of the program.

The dimensions along which these aspects of the project differ are often difficult to ascertain. Some of the failures to detect the effects of a program may be attributed to these sources of variation which may wash out effects when data are inappropriately combined for analysis. The careful examination of the characteristics of the population and sites in which the program is implemented can isolate sources of variation which hamper the detection of effects, be they negative or positive. Since we have discussed numerous sources of systematic variation encountered in our evaluation of HSPCLP in another paper (Cordray & Staneski, 1975) the present discussion will be limited to the problems associated with attrition and differential exposure to the program. These, it can be argued, are problems which are not unique to this particular program, but have implications for evaluation in any setting where a longitudinal program is implemented.

Attrition. The unexpected loss of recipients from a social program has been termed attrition or more traditionally "subject mortality" (Campbell & Stanley, 1963). This well known problem has been identified in the quasi-experimental literature as a major threat to the internal validity of a causal interpretation. Without further information regarding the differences between those who remain and those who leave (drop out) the effects attributable to the program become uninterpretable.

A related problem that is typically encountered in urban populations has been termed transiency. It could be argued that attrition is simply one aspect of this more general characteristic. High mobility (transiency) is a pervasive characteristic of urban populations and presents a particularly large problem for longitudinal evaluation. Specifically, transiency

refers to the influx and outflow of individuals in a given time frame, whereas attrition refers only to instances where the individual leaves the program before his scheduled termination. Similar to the problem of non-random assignment, one might wonder about the equivalence between those pupils who remain and those who leave. The validity of results based on a partial sample becomes a direct function of the proportion remaining at the end of the program's scheduled duration. The higher the attrition, the more questionable become results based on a low retention rate. Urban schools within the Los Angeles area consistently a transiency rate in excess of 50% per school year. If this rate were calculated over a four year period, it would be conceivable that only 1/8 of the original population would still be enrolled. Fortunately, we have found a higher degree of stability than implied above. Our records show that nearly one third of the initial Wave I population were enrolled throughout the duration of the program. To assess the validity of our results based on those who remain, it becomes necessary to determine their equivalence with those who left the project classes. High rates of transiency can affect the results in either a positive or a negative direction. In any event, the effects of transiency cannot be ascertained until systematic tests have been carried out.

Implied by the above argument are certain data gathering and research design considerations. For instance, knowledge that a program is being implemented in an area where the target population is highly mobile requires the evaluator to gather additional information on individual recipients. While the exact nature of this additional information is directed by the focus of the particular program, careful consideration

at the planning stage should be an attempt to assess factors that could potentially result in someone leaving the program. An exit interview is the most obvious means of gathering this data. Frequent monitoring of population characteristics (SES) can also provide useful information from which the plausibility of alternative explanations can be assessed. For example, if a substantial portion of the population leaves the project this may be indicative of success rather than failure, e.g., a program designed for job training may serve as a resource through which the recipients obtained jobs having not gone through the entire training process as planned. This "spin off" effect would not be detected unless the evaluation design made provisions to assess early termination.

Another consideration deals with the process of data collection and storage. Longitudinal evaluations, where transiency is a potent dimension, requires constant updating and verification of recipient characteristics. Archival records, interviews, and site visits can provide much of the necessary data for adequately describing population characteristics. The importance of information obtained from these sources becomes apparent, usually after the fact, when attempting to interpret a program's impact on its recipients. In a closed system, where attrition is the only consideration, early termination results in a decreased sample size, thereby limiting the degree to which fine grained analyses can be carried out and a loss of statistical power due to the decrease in degrees of freedom. When the system (e.g., public school) is open to new entrants the population composition changes thereby introducing additional sources of systematic variation.

Length of exposure to the program. Given the high transiency rate,

where individuals not only leave, but enter at different times, an additional problem is introduced; namely,the need to consider the individual's length of exposure to the program. For example, with respect to the HSPCLP differential exposure becomes especially important when using traditional outcome measures, such as scores in yearly achievement tests. If a program is, in fact, effectively producing the desired changes in its recipients as evidenced by these outcome measures it would be reasonable to assume that those who have had the longest exposure to the program would perform differentially on these measures. While this may not be a linear function of the duration of exposure, it certainly raises a question regarding when the program's effect is most influential on its recipients. Failure to consider the possibility that the target population composition, at any given point in time, varies with respect to length of exposure to the program can result in erroneous conclusions or spurious trends. In addition to testing the equivalence of those who attrite from the program it is necessary to examine the comparability of those new entrants.

An Example

To illustrate the relationship between length of exposure and transiency for each year of the project, Table 1 and Table 2 (for Wave 1 and Wave 2 respectively), are presented below. Examination of Table 1a shows that the total enrollment for each year (one through four) has remained relatively constant.

- - - - - - - - - - - - - - - -

Insert Tables 1 and 2 here

- - - - - - - - - - - - - - - -

However, inspection of Table 1b demonstrates that an accurate description

Note: These tabled values are from a preliminary analysis, updated information will be available from the authors after December, 1976.

of this yearly enrollment, especially as the project progresses in time, is dependent upon knowledge of the preschool school classification and length of exposure to the project (as characterized by the dimension of year of entrance). It can be seen that the composition of the groups changes considerably over time. Another characteristic of the population concerns year to year retention. Table 1c shows differential retention as a function of the preschool experience classification. These relationships are also demonstrated for Wave 2 classes in Tables 2a through 2c. These figures represent aggregate summaries of the characteristics within the entire HSPCLP population. As indicated earlier, the differences between sites (schools and classrooms) provide even more variability with respect to the composition of the project population. For example, some classes have remained relatively intact throughout the duration of the program and in other cases the high rate of transiency has resulted in very few pupils remaining for the duration. This sheds doubt on the comparability of classrooms in determining the overall effects of the program.

An indication of the movement in our population can also be seen in the number of recipients who have passed through the program. At the completion of the fourth year, in Wave I, approximately 650 pupils had received some portion of the HSPCLP. Wave II will have had contact with a similar number of recipients at its termination in June of this year. The fact that each wave consists of ten elementary classrooms, shows that a great deal of movement has occurred in these populations. In effect, this movement has drained the initial target population considerably.

Inspection of differential transiency within sites reveals that there are few pupils remaining who received the Foundation Head Start experience. This loss of subjects makes the evaluation of FHS on its recipients difficult in terms of statistical power.

A similar problem is encountered with respect to assessing the effects of the length of exposure that a recipient has had to the program. The ability to block along this dimension becomes increasingly problematic as those who entered the program early are replaced by those who entered late. All of the problems mentioned above are of the type which have the tendency of leading to erroneous conclusions as to the impact of the program. We are fortunate to have the incorporation of two waves in our evaluation design; one which serves as a replication of the other. An effect found in Wave I can be demonstrated to have a certain amount of stability to the extent to which it is replicated in Wave II. Further, the validity of alternative explanations discovered with the Wave I population can be tested on the second wave providing a feed forward information loop.

Another factor which we have considered involves the gathering of information regarding the reasons for which a recipient has left a program. For instance, a pupil may leave a HSPCLP classroom as a result of having been identified as gifted and therefore moved to a different class. Alternatively, a student may be detained due to lack of academic progress. Another, perhaps more subtle, example of differential transiency relates to that portion of our population which is Spanish. Many of these pupils left the HSPCLP after kindergarten to enter Catholic schools (which rarely have kindergartens). Information of this type is

needed to adequately assess program effectiveness.

This argument can be extended to the problem of those pupils who entered the program at different points along its duration. Here, it is necessary to ascertain both why a student has entered (i.e., has the student been detained?, is his entry due to general mobility?, etc.) and the level, in relation to existing recipients, at which he or she has entered. Most important, we feel is the administration of an entry level pretest to aid in the assessment of equivalence between entering, remaining, and leaving recipients. Given the nature of the above problems what initially appears to be a perfect example of a time series design rapidly degenerates into a quasi-experimental patch-up design (see Campbell & Stanley, 1963).

General Model

A general problem in the evaluation research setting is that of finding statistically reliable effects for which valid interpretations can be drawn. In any project there are multiple factors which inhibit the detection of effects and assessment of their validity, two of which we have discussed (transiency and length of exposure to the program). Our position is that this problem of detecting effects (Stage 1) supersedes that of interpretive validity (Stage 2). That is, unless an effect can be found, the problem of validity is of diminishing concern. Most important is that the present methodology must be able to "pick up" an effect if one exists. The foregoing pages have outlined some of the design strategies which can be employed to accomplish this task. The statistical aspects of this problem begin to take on increasing importance as one considers the number of comparisons necessarily made to

assess the equivalence of the varying elements of the target population. The probability of making both Type I (false positive) and Type II (false negative) errors needs constant attention. For the purpose of evaluation research, it might be appropriate, in some instances, to relax the probability of making a Type I error in order to maximize the probability of demonstrating an effect. Once an effect (e.g., differential treatment means) has been uncovered its validity can be scrutinized. Again, the replication wave proves useful for ascertaining validity and reliability. Threats to statistical validity (Cook & Campbell, 1976) become problematic due to the intimate relationship between Type I error, Type II error, power, and sample size. In anything but a fairly extensive program (in terms of sample size) statistical power may be insufficient to allow reliable results to be found. With the exception of perhaps the Year 4 population (which for Wave I is very heterogeneous) the sample size and statistical validity problems are not an overwhelming concern for the HSPCLP evaluation.

To illustrate how one might approach the problem of systematically examining and interpreting program effects, the following example is provided. The general model follows the two stage process outlined above. First, the data is examined for systematic sources of variation and once these effects have been examined the data is contrasted with comparison and/or base rate data. Then the validity of such findings are scrutinized by testing rival alternative explanations identified by the evaluator.

When considering the validity of achievement related outcome scores, an accurate description of the year to year population composition is

mandatory. As indicated in Tables 1 and 2 the population composition changes from year to year and there are two systematic sources of variation. (The Table also includes a breakdown according to differential preschool experience; however, for the purpose of this example its effects will be ignored.) If an evaluator failed to recognize that at each test administration respondents differed in their length of exposure to the program, this variation may result in a failure to detect program effects when contrasted with a comparison group. Consequently, this variation must be accounted for prior to the assessment of differential program effects. A framework for analyzing systematic variation is schematized in Figure 1.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

Inspection of Figure 1 demonstrates the operation of the two major components of transiency; the flow of new entrants into the project and attrition from the project. The various achievement tests and times at which they were administered is depicted by the letter T with different subscripts to identify each test. In order to examine the effects of these sources of variation a number of comparisons must be carried out. The necessary comparisons include an examination of the equivalence of test scores between (1) those who remain and those who leave, (2) between those who leave and those newly entering the program, and (3) between those who remain and those who have recently entered the project. As can be seen by inspecting Figure 1, as the project progresses in time we have increased the number of outcome measures. In addition to giving a more differentiated picture of program effects, the use of multiple

outcome measures provides a safeguard against the statistical error rate problem, i.e., differences are not considered too seriously unless they appear on a number of dimensions. Further, qualitative information (e.g., identification of gifted pupils, family upward mobility, or an educational handicap) provides information which aids in interpreting these differences.

Having assessed the effects of these sources of variation, the differential effects exhibited within a project can then be meaningfully compared with the results obtained from comparison groups. It is possible that transiency or systematic removal (e.g., ability grouping) of specific groups of pupils (e.g., FHS pupils) may result in no differential effects between the project and comparison groups. This could occur at any point along the project's time line. Similarly, differences may appear which indicate the program has been effective for those who have four years of exposure and no effects on those who have recently entered. This difference would not have appeared had the data been treated as a whole and not stratified along these relevant dimensions. As can be seen, this methodology not only allows the evaluator to assess positive and negative outcomes but also allows him to specify the conditions or process under which these results were obtained. When a replication sample is employed, assessment of the generalizability of the conclusions can also be considered.

A discussion of the technical issues associated with scaling problems, error rate, and power can be found in almost any intermediate statistic book (e.g., Hays, 1973). The problem of assessing the non-equivalence of comparison groups is outlined by Kenny (1975) and Horst, Tallmadge and Wood (1975). The complexity of these issues preclude any

more than a brief mention at this point. For the purpose of assessing nonequivalence of comparison groups, elaborate statistical procedures have been developed in an attempt to reduce the plausibility of statistical and subject selection artifacts (Kenny, 1975). As witnessed in Kenny's presentation, these models are mathematically very elegant; however, their successful application requires considerable forethought, in that additional information (not supplied by the mathematics) is necessary to assess the effects of threats to interval, external, statistical and construct validity described by Cook & Campbell (1976). An important consideration, that must always be kept in mind, is that no one study will ever provide enough information to completely eliminate the sources of invalidity. Our argument concerns the performing of those activities which most closely approximates a comprehensive understanding of the dynamics within a program. Some aspects of the natural environment tend to obscure program effects and those need to be examined.

Summary

In the present paper we have been stressing the need to ascertain the validity of evaluation data. This may appear to some individuals as overemphasizing the utility of objective information or outcome measures. On the contrary, this scrutinization should be carried out on all the data. We are fully aware of the need to take an eclectic approach to evaluation and this project is being evaluated as a dynamic system. The current analysis of program effects upon pupils is only one component of the entire evaluation.

The position being espoused is one which emphasizes the thoughtful

consideration of the multiple factors operative within the social program setting. It is believed that program effects are at times masked by systematic sources of variation among the various components of the program. These sources of variation are not necessarily a part of the program and tend to overwhelm the usually weak effects of a social program. Similarly, program effects may be erroneously positive due to these variations. The choice of stratification variables (believed to be the source of systematic bias) needs to be made after careful examination of the entire system. Isolation of trivial variants may simply confuse the evaluation effort, there are no clear cut criteria for selecting these factors. It seems that the evaluation of longitudinal social action programs, out of necessity, takes on an evolutionary characteristic--as new problems are identified, procedures must be devised to assess what effects these problems are likely to create. No matter what form of evaluation is carried out, the logic of eliminating plausible alterratives for observed effects can be applied.

The consequences of program evaluation should have more impact on the social system than merely specifying the success or failure of a particular program. For example, our analysis of the effects of transiency and differential exposure to the program could be seen as having implications for the planning, development, and implementation of future programs of this nature. For example, it could be argued that longitudinal programs, directed at a highly mobile target population, should not be undertaken at the very outset. It could be further argued that since it is difficult to effect social change when the target population is not stable (in terms of mobility) enough to recieve the

benefits of the program that it is bound to fail to produce its intended effects. This would be especially true when the goals are longitudinal in nature. Rather than taking this position, we would argue that these factors should be recognized and the program goals and methods for achieving those goals should be modified accordingly. The information generated from other projects will set limits upon the feasibility of a future program's goals which in turn will affect the proposed program's conceptualization of its objectives, activities, and evaluation design. In most cases of longitudinal programs the goals are not only short range or long range but there are interrelationships among goals for various components of the project. This fact will necessitate trade-offs between aspects of the programs in terms of the feasibility of various objectives. For example, transiency in the HSPCLP primarily affects the goals associated with the pupil population, it does not affect those objectives associated with the Continuation Teachers who are served by the project. Thus, a decision to discontinue these types of projects because of pupil mobility would interfere with the other aspects of the project. Instead one might reconsider the role of pupil mobility in terms of the function of the Continuation Teachers and devise an alternative set of objectives which would maximize shorter term effects for the pupils and allow the long term objectives a chance to operate for the other components of the project (career development for the staff and continuation teachers).

The information generated by a careful examination of the system imposed constraints on a social action program should provide valuable information that can be used to select among the most feasible potential intervention strategies.

REFERENCES

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental design for research. Chicago: Rand McNally, 1963

Cook, T. D. & Campbell, D. T. The design of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), Handbook of industrial and organizational research. Chicago: Rand McNally, 1976. In press.

Cordray, D. S. & Staneski, R. A. Evaluation strategies for longitudinal urban educational programs: A case history. Paper presented at the First National Conference on Urban Education, Kansas City, Mo., November 1975.

Hays, W. L. Statistics for the social sciences (second edition) New York: Holt-Rinehart & Winston, 1973.

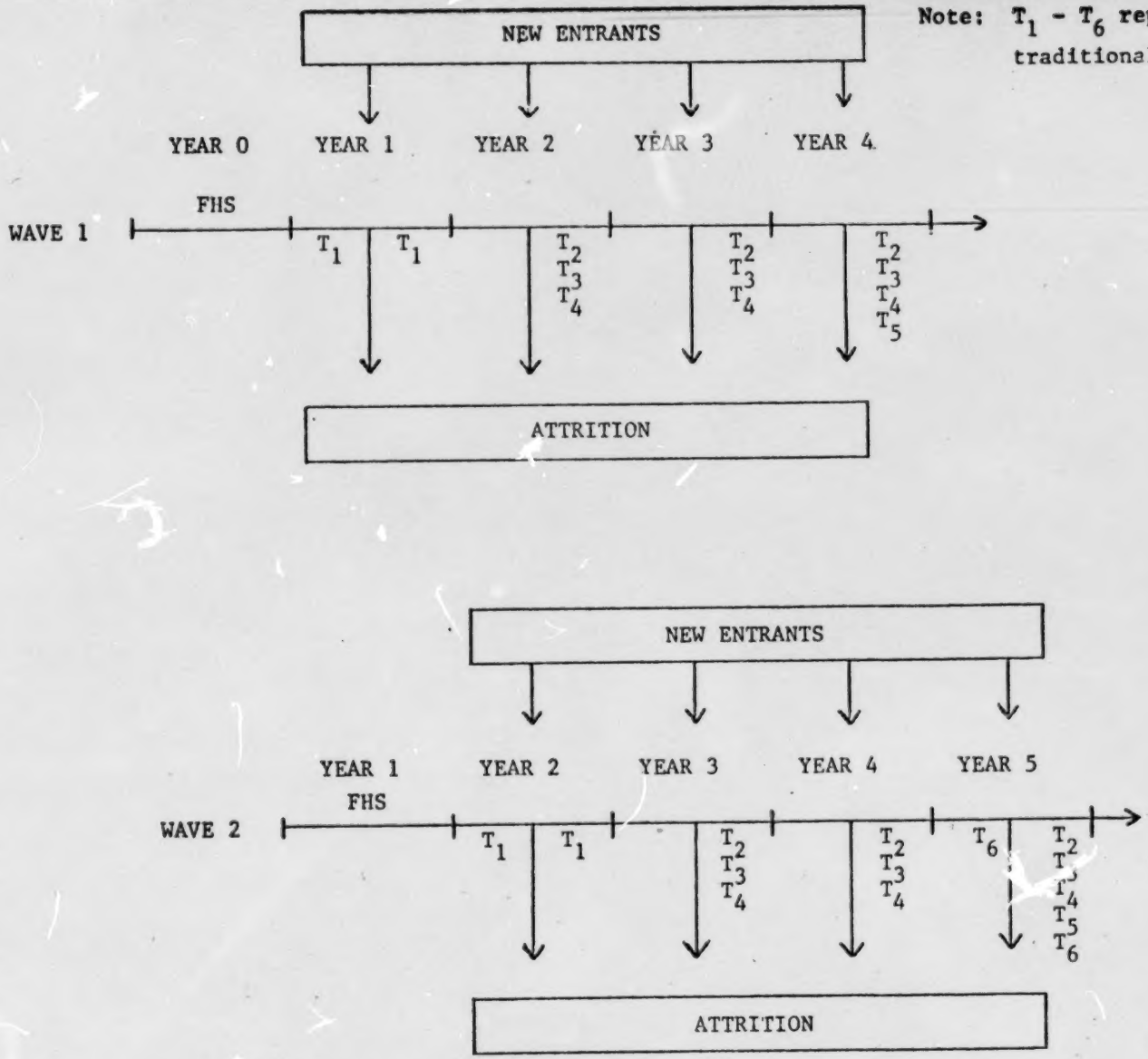
Horst, D. P., Tallmadge, G. K., & Wood, C. T. A practical guide to measuring project impact on student achievement No. 1. U. S. Department of Health, Education and Welfare (Document No. 7M004 048), 1975

Hyman, H. H. & Wright, C. R. Evaluating social action programs. In P. F. Lazarsfeld, W. H. Sewell, and H. L. Wilensky (Eds.), The uses of sociology. Basic Books, 1967, pp. 741-782.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in non-equivalent control group design. Psychological Bulletin, 1975 82, 345-362

Tripodi, T., Fellin, P., & Epstein, I. Social program evaluation. Itasca, Ill.: Peacock, 1971

Weiss, C. Evaluation research. New York: Prentice Hall, 1972



Note: T_1 - T_6 represent scores on traditional outcome measures.

Fig. 1. RELATIONSHIP BETWEEN TRANSIENCY AND VALIDATION OF OUTCOME MEASURES.

HEAD START PRIMARY CONTINUATION LEARNING PROJECT: WAVE I

TABLE 1a

YEARLY ENROLLMENT (TOTAL) AND YEARLY ENROLLMENT ACCORDING TO DIFFERENT CLASSIFICATIONS OF PRESCHOOL EXPERIENCE

PROJECT YEAR	PRESCHOOL EXPERIENCE	ENROLLMENT	PERCENTAGE OF YEARLY ENROLLMENT
YEAR 1	FHS	105	37.8%
	OTHER PS & HS	50	18.5%
	NO PRESCHOOL	116	42.8%
	TOTAL	271	100 %
YEAR 2	FHS	73	29.1%
	OTHER PS & HS	37	14.7%
	NO PRESCHOOL	141	56.2%
	TOTAL	251	100%
YEAR 3	FHS	72	24.2%
	OTHER PS & HS	47	15.8%
	NO PRESCHOOL	179	60.0%
	TOTAL	298	100%
YEAR 4	FHS	59	21.4%
	OTHER PS & HS	61	22.1%
	NO PRESCHOOL	156	56.5%
	TOTAL	276	100%

TABLE 1b

COMPOSITION OF THE POPULATION FOR EACH PROJECT YEAR ACCORDING TO PRESCHOOL EXPERIENCE AND YEAR OF ENTRANCE

PROJECT YEAR	PRESCHOOL EXPERIENCE	YEAR OF ENTRANCE YEAR 1	YEAR 2	YEAR 3	YEAR 4
YEAR 1	FHS	105 (37.9%)			
	OTHER PS & HS	50 (18.5%)			
	NO PRESCHOOL	116 (42.8%)			
	TOTAL	271 (100%)			
YEAR 2	FHS	65 (25.9%)	8 (3.2%)		
	OTHER PS & HS	22 (8.8%)	15 (6.0%)		
	NO PRESCHOOL	53 (21.1%)	88 (35.1%)		
	TOTAL	140 (55.8%)	111 (44.2%)		
YEAR 3	FHS	60 (20.1%)	6 (2.0%)	6 (2.0%)	
	OTHER PS & HS	19 (6.4%)	10 (3.4%)	18 (6.0%)	
	NO PRESCHOOL	35 (11.7%)	51 (17.1%)	93 (31.2%)	
	TOTAL	114 (38.2%)	67 (22.5%)	117 (39.3%)	
YEAR 4	FHS	47 (17.0%)	2 (.7%)	6 (2.2%)	4 (1.4%)
	OTHER PS & HS	17 (6.1%)	8 (2.9%)	15 (5.5%)	21 (7.6%)
	NO PRESCHOOL	25 (9.1%)	32 (11.6%)	37 (13.4%)	62 (22.5%)
	TOTAL	89 (32.2%)	42 (15.2%)	58 (21.1%)	87 (31.5%)

TABLE 1c

RETENTION RATES* FOR EACH CLASSIFICATION OF PRESCHOOL EXPERIENCE ACCORDING TO PROJECT YEAR AND YEAR OF ENTRANCE

PRESCHOOL EXPERIENCE	PROJECT YEAR	YEAR OF ENTRANCE YEAR 1	YEAR 2	YEAR 3
FHS	YEAR 2	61.9%		
	YEAR 3	57.1%	75%	
	YEAR 4	44.8%	25%	100%
OTHER PS & HS	YEAR 2	45.5%		
	YEAR 3	38.0%	67%	
	YEAR 4	34.8%	53%	83.3%
NO PRESCHOOL	YEAR 2	45.7%		
	YEAR 3	30.2%	57.9%	
	YEAR 4	21.6%	36.8%	47.5%
AVERAGE RETENTION	YEAR 2	51.7%		
	YEAR 3	42.1%	60.4%	
	YEAR 4	32.8%	37.8%	50.5%

* (expressed as a percentage of those remaining between years)

HEAD START PRIMARY CONTINUATION LEARNING PROJECT: WAVE II

TABLE 2a

YEARLY ENROLLMENT (TOTAL) AND YEARLY ENROLLMENT ACCORDING TO DIFFERENT CLASSIFICATIONS OF PRESCHOOL EXPERIENCE

PROJECT YEAR	PRESCHOOL EXPERIENCE	ENROLLMENT	PERCENTAGE OF YEARLY ENROLLMENT
YEAR 2	FHS	154	48.9%
	OTHER PS & HS	46	14.6%
	NO PRESCHOOL	115	36.5%
	TOTAL	315	100%
YEAR 3	FHS	109	37.8%
	OTHER PS & HS	45	15.6%
	NO PRESCHOOL	134	46.5%
	TOTAL	288	100%
YEAR 4	FHS	88	30.8%
	OTHER PS & HS	46	16.1%
	NO PRESCHOOL	152	53.1%
	TOTAL	286	100%

TABLE 2b

COMPOSITION OF THE POPULATION FOR EACH PROJECT YEAR ACCORDING TO PRESCHOOL EXPERIENCE AND YEAR OF ENTRANCE

PROJECT YEAR	PRESCHOOL EXPERIENCE	YEAR OF ENTRANCE YEAR 2	YEAR 3	YEAR 4
YEAR 2	FHS	154 (48.9%)		
	OTHER PS & HS	46 (14.6%)		
	NO PRESCHOOL	115 (36.5%)		
	TOTAL	315 (100%)		
YEAR 3	FHS	106 (36.8%)	3 (1.0%)	
	OTHER PS & HS	27 (9.4%)	18 (6.3%)	
	NO PRESCHOOL	54 (18.8%)	80 (27.6%)	
	TOTAL	187 (65.0%)	101 (35.0%)	
YEAR 4	FHS	81 (28.3%)	3 (1.0%)	4 (1.4%)
	OTHER PS & HS	15 (5.3%)	9 (3.1%)	22 (7.7%)
	NO PRESCHOOL	30 (10.5%)	38 (13.3%)	84 (29.4%)
	TOTAL	126 (44.1%)	50 (17.4%)	110 (38.5%)

TABLE 2c

RETENTION RATES* FOR EACH CLASSIFICATION OF PRESCHOOL EXPERIENCE ACCORDING TO PROJECT YEAR AND YEAR OF ENTRANCE

PRESCHOOL EXPERIENCE	PROJECT YEAR	YEAR OF ENTRANCE YEAR 2	YEAR 3
FHS	YEAR 3	68.8%	100%
	YEAR 4	52.6%	
OTHER PS & HS	YEAR 3	58.7%	50%
	YEAR 4	32.6%	
NO PRESCHOOL	YEAR 3	46.9%	47.5%
	YEAR 4	26.1%	
TOTAL	YEAR 3	59.4%	49.5%
	YEAR 4	40.0%	

* (expressed as a percentage of those remaining between years)